

N00014-95-1-1088 (DTIC)

A Preliminary Model of Centering in Dialog

D.K. Byron and A.J. Stent

Technical Report 687
September 1998

19990423 004

DTIC QUALITY INSPECTED 2

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

UNIVERSITY OF
ROCHESTER
COMPUTER SCIENCE

A Preliminary Model of Centering in Dialog*

D. Byron and A. Stent
Department of Computer Science
University of Rochester
Rochester NY 14627, U.S.A.
dbyron/stent@cs.rochester.edu

The University of Rochester
Computer Science Department
Rochester, New York 14627

Technical Report 687

September 25, 1998

Abstract

The centering framework explains local discourse coherence by relating a speaker's focus of attention and the forms of referring expressions. Although this framework has proven useful in single-speaker discourse, its utility for multi-party discourse has not been shown. It is unclear how to adapt it to handle discourse phenomena such as turn-taking, acknowledgments, first and second person pronouns, and disfluencies. This paper reports our experiments applying three naive models of centering theory for dialog. These results will be used as a baseline for future, more sophisticated models.

This material is based on work supported by NSF grant IRI-96-23665, ONR grant N00014-95-1-1088 and Columbia University grant OPG:1307. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the granting institutions.

A shorter version of this paper appeared in the Proceedings of ACL-98. The authors would like to thank James Allen, Marilyn Walker, and the anonymous reviewers for their helpful comments on a preliminary draft of the paper.

REPORT DOCUMENTATION PAGE

Form Approved

OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE September 1998	3. REPORT TYPE AND DATES COVERED technical report
4. TITLE AND SUBTITLE A Preliminary Model of Centering in Dialog		5. FUNDING NUMBERS ONR N00014-95-I-1088
6. AUTHOR(S) Donna K. Byron and Amanda J. Stent		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESSES Computer Science Dept. 734 Computer Studies Bldg. University of Rochester Rochester NY 14627-0226		8. PERFORMING ORGANIZATION
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of Naval Research Information Systems Arlington VA 22217		10. SPONSORING / MONITORING AGENCY REPORT NUMBER TR 687
11. SUPPLEMENTARY NOTES		
12a. DISTRIBUTION / AVAILABILITY STATEMENT Distribution of this document is unlimited.		12b. DISTRIBUTION CODE
13. ABSTRACT (Maximum 200 words) (see title page)		
14. SUBJECT TERMS centering; dialog; discourse modelling		15. NUMBER OF PAGES 8 pages
		16. PRICE CODE free to sponsors; else \$2.00
17. SECURITY CLASSIFICATION OF REPORT unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT unclassified
		20. LIMITATION OF ABSTRACT UL

1 Introduction

One of the most influential models in computational linguistics relating a speaker's local focus of attention to the form chosen for referring expressions is the centering framework [3; 4]. Although it was originally intended to explain local coherence, later research [1] showed that it could also be used to bind ambiguous pronouns. A number of subsequent studies have explored refinements and extensions of the theory (eg. [10; 13; 8; 14]), but only a few have attempted to extend the centering framework from single-speaker to multi-party discourse (cf. [2; 9; 15]).

If the claims of centering theory hold true for multi-party discourse, it could be of use in both generation and understanding components of natural language dialog systems. Using the centering framework, dialog understanding systems would have a more accurate representation of attentional state, and therefore more accurate reference resolution. Dialog generation systems would have a better model of local focus, resulting in more coherent referring expressions. The choice of referring expression in generated discourse is important, since it has been shown that infelicitous use of a pronoun or definite noun phrase when the other should be used increases the listener's processing time [7].

Although the benefits of centering theory for dialog systems are clear, it is not at all clear how to adapt it to multi-party discourse. A primary use of centering is its predictions about how the current utterance will affect the next; so one might ask whether it is even appropriate to use the model when successive utterances might not be produced by the same speaker. Will a different speaker focus on the same entity the original speaker had in mind, and if so will the form of referring expressions follow the same patterns between speakers?

This paper describes three alternative experimental models of the centering framework, evaluated on a corpus of 2-person social conversations. We chose very naive approximations to the original framework as a starting point. Their performance will be used as a baseline for evaluating more sophisticated models in the future.

We first give a brief summary of the original definition of centering theory, then describe some of the particular issues that create a challenge for adapting it to spoken dialog. Finally, we present our three alternative formulations of the theory and discuss the results of applying them to a corpus of 2-person social conversations.

2 The Centering framework

The centering framework makes three main claims: 1) given an utterance U_n , it predicts which discourse entity will most likely be the focus of the next utterance U_{n+1} ; 2) when the local focus is maintained from one utterance to the next, the framework predicts that it will be expressed with a pronoun; and 3) when a pronoun is encountered, centering provides a partial ordering on possible antecedents from the prior utterance.

In the centering framework, the following data structures are constructed for each utterance U_n :¹

1. a partially-ordered list of forward-looking centers Cf_n that includes all the discourse entities in the sentence. The first element of this list is called the Cp_n (preferred center).
2. a backward-looking center Cb_n that is the most highly ranked element of Cf_{n-1} that appears in Cf_n . The Cb_n is the central entity, or 'center', of U_n .

¹We provide only the briefest sketch of the centering framework. Readers unfamiliar with this theory are referred to [4].

$Cb_n = Cb_{n-1}$	$Cb_n \neq Cb_{n-1}$	
$Cb_n = Cp_n$	Continuing	Shift
$Cb_n \neq Cp_n$	Retaining	Shift

Table 1: Transition states in the Centering framework

Two crucial factors that impact the implementation of centering theory are the selection of items for Cf and their rank order in the list (because that determines the Cb). All discourse entities ‘realized in the utterance’ should be included in Cf , but depending on the semantic model used, this might include only entities explicitly expressed in the utterance or additional associated entities (such as part/subpart and subtype/supertype) and ellipsed elements. Although the ordering for the Cf list is an active area of research, we adopt the following partial ordering for the purposes of this paper: Subject > Direct Object > Indirect Object > Complements > Adjuncts [1; 8].² Linear order in the sentence is used to complete the ordering.

Centering theory defines a preference ordering on techniques for effecting a topic change, as shown in Table 1. Transition types are ‘shift’, ‘retain’ and ‘continue’, based on whether $Cb_n = Cb_{n-1}$ and whether $Cb_n = Cp_n$.

At the heart of the theory are the two *centering rules*:

Rule 1: If any member of Cf_n is realized by a pronoun in Cf_{n+1} , Cb_{n+1} must also be a pronoun.
Rule 2: Sequences of continues are preferred over sequences of retains, and sequences of retains are preferred over sequences of shifts.

Although the centering framework was originally intended to explain local coherence, subsequent research by [1] showed that it could also be used to bind ambiguous pronouns. Their algorithm used the preference ordering of transition types described by centering rule 2 to choose between competing referents for ambiguous pronouns. Because the Cf is represented as a list instead of as one item, the authors claim that the centering model can dismiss with the multiple categories of focus used in Sidner’s model [12]. Centering also handles some cases of multiple pronouns that were problematic in Sidner’s model by allowing other entities to be pronominalized as long as the Cb is [3].

The collection of papers from the Penn centering workshop offers a variety of new studies to deal with limitations in the original centering framework and investigations into its utility for research in languages other than English [14]. Brennan’s paper describes an initial attempt to address the problems of applying centering to dialog, including many of the issues addressed in this paper [2].

3 Issues in adapting centering for two-party dialog

A variety of issues must be addressed to adapt centering to two-party dialog. Among them are:

- Utterance boundaries are difficult to pin down in spoken dialog, and their determination affects the Cf lists. Just how the speaker turns are broken into utterances has a huge impact on the success of the theory in explaining coherence and binding pronouns [2].
- The subject of the sentence is ranked first in Cf , however in dialog the subject is often a first or second person pronoun. Should these pronouns be included in Cf ?

²Subsequent work [8] suggests that topicality should dominate over sentential role in determining the ordering of Cf .

- Which utterance should be considered the ‘previous utterance’ for the purpose of locating the C_b candidate set - the same speaker’s previous utterance or the immediately preceding utterance, regardless of speaker?
- What should be done with abandoned or partial utterances and utterances with no discourse entities (such as acknowledgments)?

Many other issues, including discourse phenomena such as grounding and adjacency pairs, could also be considered, but are beyond the scope of this preliminary study.

4 Experimental method

Our data come from four randomly selected dialogs from the CALLHOME-English corpus³ [11]. One dialog was used to train the annotators (the authors). Because annotators must not confer during the process of annotation, a training dialog is often used to clarify policies for borderline cases for which the annotation instructions are unclear. In our case, we established policies on the types of surface forms that would be considered to represent discourse entities and the criteria for breaking a discourse into non-subordinate clauses. After training, the other three dialogs were independently annotated according to the rules outlined below.

The annotators compared their results and agreed upon a reconciled version of the data. The reconciled version was used to produce the results reported in Section 5. Annotator accuracy as measured against the reconciled data over all categories (C_b model 1, C_b model 2, C_b model 3, and C_f) ranged from 79.8% to 89.2%. Accuracy was calculated by counting the number of utterances with one or more differences from the reconciled data (including ordering differences in the case of C_f lists), and dividing by the total number of utterances.

More standard reliability measures could not be used as there are no “tags” in this annotation scheme, and within some categories there may be an ordered list of items.

Discourse segmentation was done by the annotators jointly before annotation took place, using criteria similar to those laid out in [6]; the purpose of this study was not to examine means of segmenting discourse, but extensions of centering theory.

We honored all utterance boundaries in the original transcripts⁴, even if an utterance appeared to be a fragment properly belonging to a preceding or following utterance. For instance, each of the following two-utterance pairs seem as though it should be just one utterance:

Example 1 [dialog 4571]

utt 443.00 A ... and she called me one day when
 utt 457.01 A there was nobody in the house but her and I drove ...

Example 2 [dialog 4861]

utt 519.54 A I know there’s a limit for
 utt 521.25 A students who e-

Example 3 [dialog 4861]

utt 1075.00 B The only people who really think they know something
 utt 1078.19 B are the people who usually are deficient.

³The dialog transcripts consisted of 614 utterances, 6628 words, 30 minutes of speech. After annotation, there were 864 utterances, of which 200 were empty (contained no discourse entities).

⁴Transcribers of the CALLHOME corpus were instructed to separate utterances if there was a speaker change, the semantics or syntax of the language indicated the end of an utterance, or there was a long pause.

For utterances consisting of compound sentences, we added utterance breaks between non-subordinate clauses. The utterance breaks we added in the following sentences (each originally transcribed as a single utterance) are indicated with /:

Example 4 [dialog 4248]

"It does make a difference / like I always thought formula smells kind of disgusting."

Example 5 [dialog 4571]

"...in an apartment and / that's going to be really, I hope it's going to be nice. / His buddy is here actually and / Oscar is finishing some work that he was doing and / then he's coming up sometime next week {breath}"

For the rest of this paper, we will use the term "utterance" to indicate a non-subordinate clause.

Elements of each utterance's *Cf* were created by nouns in the utterance. Although the centering framework leaves open the question of what types of syntactic constituent contribute semantic entities to the utterance, we chose to allow only nouns to cause an element to be added to the utterance's *Cf*. We did not include other pro-forms such as temporals or complex nominal constituents such as infinitival phrases. *Cf*'s were ordered according to the syntactic model discussed in Section 2.

The 'real topic' of each utterance was selected from its *Cf* according to the annotator's intuition of what the utterance was 'about'. It had to be an entity explicitly referred to in the utterance, but could be, for instance, an entity referred to using a first or second person pronoun or something other than the *Cp* of the utterance. See Section 5 for examples.

Because originally the centering framework was proposed to handle third-person pronouns and definite NP's, most studies have dealt with text or narrations wherein referring expressions about the creator of the discourse or its addressee(s) are not included. But in some kinds of dialog, first and second person pronouns, referring to the dialog participants, abound. These expressions are very perplexing from a centering viewpoint. The speaker has no other felicitous choice of referring expression to refer to himself, so the choice of referring expression cannot be said to reflect his attentional state and we do not need any information from a centering model to aid in generating these expressions. If we are using centering to disambiguate pronouns, we should not include them because they typically do not need to be disambiguated (except in quoted contexts). However, if we claim that centering captures "what a discourse most centrally concerns"[4] and represents the local focus of attention, this role can legitimately be held by a dialog participant. For example, the following utterance:

Example 6 [dialog 4248]

"Well Norm and I are worse because we live in the suburbs."

seems to be 'about' the entities referred to by "Norm and I" so it does seem that they should be included in that utterance's *Cf*. It is often difficult to determine whether the utterance is actually about the person so described or whether he is simply the primary actor in the event being described. Furthermore, it seems possible that disregarding first and second person pronouns can affect the determination of discourse coherence. Therefore we were unsure about how to annotate first and second person pronouns in the dialogs. Our results differentiate between a model in which entities referred to by first and second person pronouns were included and two models in which they were not.

Associations (eg. specific/generic relations) and ellipsis were not allowed in determining the discourse entities. For example:

Example 7 [dialog 4248]

"Show off *le bebe*": the ellipsed 'you' was not considered.

Example 8 [dialog 4248]

"After your baby drinks the milk yeah": the bottle containing the milk was not considered

In determining the previous utterance, intervening “empty utterances” were skipped [9; 15]. An utterance was considered empty if it contained no discourse entities. Empty utterances include statements like “yeah”, “really”, as well as utterances like “hard to leave behind” (which was a complete speaker turn) that have no explicitly mentioned objects.

4.1 The proposed models

After examining the kinds of phenomena that occur in non-task-oriented dialog (and before dialog annotation), we formulated three slightly different models for determining the Cb , staying as close as possible to the original centering model.

Model 1: Cf^5 can contain entities referred to using first and second person pronouns, and Cb_n can be such an entity if it is the highest ranked element of Cf_{n-1} that appears in Cf_n . Only Cf_{n-1} , regardless of its speaker, is searched for Cb_n .

Models 2 and 3 do not include first and second person pronouns in Cf .

Model 2: Cb_n is the highest ranked item in Cf_n that occurs in either the *current* speaker’s most recent prior utterance or the *other* speaker’s most recent prior utterance. Recency takes precedence, so if an element of the Cf of the most recent utterance (regardless of speaker) appears in Cf_n , it is considered to be the Cb_n , even if a higher ranked element of the Cf list of the other prior utterance is in Cf_n .

Model 3: Only Cf_{n-1} , regardless of its speaker, is searched for Cb_n .

5 Results and analysis

Centering Rule 1 was violated in only 10 of 664 utterances, so centering theory’s assumptions linking local focus to pronouns appear to hold in dialog. We did not find speaker change to impact the center in any measurable way. This section discusses our findings, summarized in Table 2. It is interesting to note that model 2, in which we allowed the Cb to come from either the current speaker’s or the other speaker’s prior turn, performed slightly better in all categories than model 3, in which the Cb could only come from the immediately previous utterance. Model 1, which includes first and second person pronouns, had more cheap transitions than either of the other two models, so may better reflect discourse coherence.

5.1 Percent of empty Cb ’s

As shown in first section of Table 2, each of our models left at least 52% of utterances with no prediction of the Cb (no element of Cf_n is in Cf_{n-1}). The worst case, model 3, found a Cb only 27% of the time. 57% of the Cb ’s in model 1 are entities referred to using first or second person pronouns⁶.

Some empty Cb ’s occur when the two speakers made related, but topically disjoint, contributions, as in (point of interest indicated with *):

⁵The dialogs were annotated for global discourse structure jointly by the authors, so U_{n-1} is the previous utterance within the discourse segment, not necessarily the previous utterance in linear order. However, only rarely did this affect the determination of the Cb .

⁶If first and second person pronouns are disallowed, model 1 becomes equivalent to model 3.

	empty C_b			$C_b = \text{topic}$			cheap transitions			expensive trans.		
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
Dialog 1: 227 utts	110	156	169	71	49	47	94	48	47	133	144	145
Dialog 2: 229 utts	105	174	176	87	41	38	93	37	37	136	149	149
Dialog 3: 208 utts	103	137	139	77	54	54	84	58	58	114	123	123
Σ for all dialogs	318	467	484	235	144	139	271	143	142	383	416	417
Model total / 664 total utts			transition type / total transitions									
48%	70%	73%	35%	22%	21%	41%	26%	25%	59%	74%	75%	

Table 2: Comparison of three alternative centering models for dialog

Example 9 [dialog 4248]

B {inhale} And they don't need that many clothes when they're that young. And they don't need toys.
B So.
B *You know basically you don't need to bring anything.
B {lipsmack} But %uh but you know your two boobs.
{laugh} and %uh {sniff}
A {laugh}
B And since I can't really go anywhere without those anyhow {laugh}

Others result from abrupt shifts to a new topic, for instance:

Example 10 [dialog 4248]

B {inhale} And now you guys have a five year old and
B {laugh} yeah.
A *This is kind of cute. / Oh look there's the Benneton outlet.

In many cases, a C_b would have been found if we had used a more sophisticated semantic model of discourse entities, including associated and ellipsed entities in C_f . In example 11, if we had included the ellipsed location in A's utterance, it would have been the center:

Example 11 [dialog 4248]

B ... I've been there wait, yes three times I think
A Well this is our second time

5.2 C_b Matches the ‘real topic’

For utterances where a C_b was found, the C_b matched the ‘real topic’ only 21% to 35% of the time. By this measure, our models are poor predictors of local focus. Sometimes there is a sequence of utterances having the same topic, and in that case the topic of the first sentence is not necessarily predicted by centering theory. In other cases, the ‘real topic’ is some element in the C_f list other than the C_p .

For instance, in example 12:

Example 12 [dialog 4248]

A And like we went into Jackson, the town and we were like –
A – %ah let me {laugh} out of here

the ‘real topic’ for both utterances is Jackson, but according to model one the entities referred to by “we” are the C_b ’s of both utterances.

In the following, according to models 2 and 3 the Cb of both utterances is the entity referred to using "she" and "her", but according to model 1 the entity referred to with "I" is the Cb of the second utterance (annotated utterance breaks have been marked with /):

Example 13 [dialog 4861]

A saying I should call her about openings in the district / that she called me because

Finally, in example 14 the 'real topic' of the first utterance is Mama Sadie, and of the second, the entities referred to by "everybody". However, all three models have an empty Cb for the first utterance, and the entity referred to by "it" as the Cb for the second:

Example 14 [dialog 4861]

B I remember Mama Sadie was upset about it. / But everybody knew why I was doing it ...

In many cases, it seemed to us that there was more than one entity that was the topic of an utterance. Once we began annotating, we realized that there was some confusion in determining the 'real topic' of the utterance. It might be the local or the global discourse topic or it might be the central actor of the event described in the utterance, especially if the utterance involves (self-)reporting.

For instance, in example 15 it seems that the event and the central actor are equally topical in the utterance.

Example 15 [dialog 4861]

And I went across the river that day

5.3 Cheap versus expensive transitions

Strube and Hahn [13] propose a method of evaluating a model against centering rule 2, measuring the 'cost' of the listener's inference load. A transition is cheap if $Cb_n = Cp_{n-1}$, otherwise it is expensive. Models with a high percentage of cheap transitions better reflect human notions of coherence. A preferred model is one that produces a high ratio of cheap transitions. All three of our models produced a very low percent of cheap transitions in this experiment, especially when compared to Strube and Hahn's result of 80% cheap transitions.

6 Summary

We conclude that centering behaviour in dialog is consistent with that found in single-speaker discourse; however, the utility of these preliminary models is questionable. We believe that by revising our formulation of Model 1, a reliable and useful model of centering in dialog can be built.

7 Future work

This preliminary study indicates a variety of promising directions for future research. Some that we intend to pursue are:

- Evaluate the models using other criteria, e.g. the increase in amount of third person pronoun resolution.
- Examine the utility of modifying the Cf ordering as suggested by [13].
- Modify utterance boundaries to re-attach interrupted utterances.

- Re-segment the dialogs using theories such as that developed by Kameyama [14] and then examine the interactions between centering and global discourse structure.
- Improve our semantic theory so Cf lists can include associated and ellipsed entities and personal pronouns.
- Re-annotate the utterances replacing the ‘real topic’ with separate ‘actor’ and ‘discourse’ foci, then see which one the centering models most closely track.
- Examine interactions between local and global focus and notions of ‘topic’.

References

- [1] Brennan, Friedman, and Pollard. A centering approach to pronouns. In *Proceedings of ACL '87*, 1987.
- [2] Susan E. Brennan. Centering as a psychological resource for achieving joint reference in spontaneous discourse. In [14], 1998.
- [3] Grosz, Joshi, and Weinstein. Providing a unified account of definite noun phrases in discourse. In *Proceedings of ACL '83*, pages 44–50, 1983.
- [4] Grosz, Joshi, and Weinstein. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2), 1995.
- [5] Barbara J. Grosz, Karen Sparck Jones, and Bonnie Lynn Webber, editors. *Readings in Natural Language Processing*. Morgan Kaufmann Publishers, 1986.
- [6] Barbara J. Grosz and Candace L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, 1986.
- [7] R. Guindon. Anaphora resolution: short term memory and focusing. In *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics (ACL '85)*, pages 218–227, 1985.
- [8] U. Hahn and M. Strube. Centering in-the-large: Computing referential discourse segments. In *Proceedings of ACL '97*, 1997.
- [9] F. Hurewitz. A quantitative look at discourse coherence. In [14], 1998.
- [10] Megumi Kameyama. A property-sharing constraint in centering. In *Proceedings of ACL '86*, 1986.
- [11] *CALLHOME American English Speech*. Linguistics Data Consortium, 1997.
- [12] Candace L. Sidner. Focusing in the comprehension of definite anaphora. In Michael Brady and Robert C. Berwick, editors, *Computational Models of Discourse*, pages 363–394. MIT Press, Cambridge, 1983. Reprinted in [5].
- [13] Michael Strube and Udo Hahn. Functional centering. In *Proceedings of ACL '96*, 1996.
- [14] Walker, Joshi, and Prince, editors. *Centering Theory in Discourse*. Clarendon Press, Oxford, 1998.
- [15] Marilyn A. Walker. Centering, anaphora resolution, and discourse structure. In [14], 1998.